AMENDMENTS TO THE CLAIMS:

This listing of claims will replace all prior versions, and listings, of claims in the application:

LISTING OF CLAIMS:

1. (ORIGINAL) A method of displaying files within a file system to a user in a semantic hierarchy, the method comprising the steps of:

mapping the files into a semantic vector space;

clustering the files within said space; and

displaying the files in a hierarchical format based on the resulting clusters.

2. (ORIGINAL) The method according to claim 1, wherein the step of clustering the files is performed as a background routine during the operation of a computer associated with said file system.

3. (ORIGINAL) The method according to claim 2, wherein the step of clustering the files is performed in response to the creation of a new file within the file system.

4. (ORIGINAL) The method according to claim 1, wherein said files are text documents and said mapping is conducted on the basis of a language model.

5. (ORIGINAL) The method according to claim 4, wherein said mapping step comprises the steps of constructing a matrix which associates each word in the documents with a vector and associates each document with a vector.

6. (ORIGINAL) The method of claim 5, further including the step of decomposing said matrix to define the words and documents as vectors in a continuous vector space.

7. (ORIGINAL) The method of claim 5, wherein said clustering is performed by identifying documents whose vectors are within a threshold distance of one another.

8. (ORIGINAL) The method of claim 7, further including the step of defining multiple threshold values and clustering said documents in accordance with said multiple threshold values to thereby establish plural levels of clusters.

9. (CURRENTLY AMENDED) The method of claim 5 further including the step of automatically labeling the clusters based on the resulting clusters.

10. (ORIGINAL) The method of claim 9 wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

11. (CURRENTLY AMENDED) A graphical user interface configured to display files in a virtual file system with a semantic hierarchy, wherein the semantic hierarchy is based on clustering of files based on semantic similarities.

12. (CANCELED)

13. (CURRENTLY AMENDED) The graphical user interface according to claim ~~12~~ 11, wherein clustering of the files is initiated by user selection.

14. (CURRENTLY AMENDED) The graphical user interface according to claim ~~12~~ 11, wherein clustering of the files is initiated upon creation of a new file in the file system.

15. (CURRENTLY AMENDED) The graphical user interface according to claim ~~12~~ 11, wherein text files are clustered utilizing a language model and non-text files are clustered utilizing rule-based techniques.

16. (ORIGINAL) The graphical user interface according to claim 15, wherein said language model comprises the LSA paradigm.

17. (CURRENTLY AMENDED) Computer readable media having stored therein computer executable code for analyzing files in a file system to determine similarities in data pertaining to their content, determining a directory structure based

on determined similarities between the files, and displaying files in hierarchical format based on the determined similarities between the files.

18. (ORIGINAL) The computer-readable media of claim 17 wherein said files are text documents, and the similarities are based upon the word content of the files.

19. (ORIGINAL) The computer-readable media of claim 18 wherein said similarities are determined in accordance with a language model, and the files are clustered in accordance with said model.

20. (ORIGINAL) The computer-readable media of claim 19, wherein said language model comprises the LSA paradigm.

21. (CURRENTLY AMENDED) The computer-readable media of claim 19, wherein said computer-executable code performs the steps of constructing a matrix which associates each word in the documents with a vector and associates each document with a vector.

22. (ORIGINAL) The computer-readable media of claim 21, wherein said computer-executable code further performs step of decomposing said matrix to define the words and documents as vectors in a continuous vector space.

23. (ORIGINAL) The computer-readable media of claim 22, wherein said computer-executable code performs clustering by identifying documents whose vectors are within a threshold distance of one another.

24. (ORIGINAL) The computer-readable media of claim 23, wherein said computer-executable code further performs step of clustering said documents in accordance with multiple threshold values to thereby establish plural levels of clusters.

25. (CURRENTLY AMENDED) The computer-readable media of claim 19, wherein said computer-executable code performs step of automatically labeling the clusters based on the resulting clusters.

26. (ORIGINAL) The computer-readable media of claim 25, wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

27. (CURRENTLY AMENDED) The computer readable media according to claim ~~16~~ 17, wherein the computer executable code performs the following steps:

clustering text files within the file system using semantic similarities;

clustering non-text files within the files system using rule-based techniques;

labeling the resulting clusters; and

displaying the files in a hierarchical format based on the resulting clusters and labels.

28. (CURRENTLY AMENDED) A computer system, comprising:

a file system storing files;

a display device; ~~and~~

a processor for analyzing the content of files stored in said file system to map said files into a semantic vector space and cluster the files within said space;

a user interface which displays representations of files stored in said file system in the form of a semantic hierarchy that is based upon the content of said files, wherein said user interface displays said files in accordance with said clustering.

29. (CANCELED)

30. (CURRENTLY AMENDED) The computer system of claim ~~29~~ 28, wherein said files are text documents and said processor maps said files on the basis of a language model.

31. (ORIGINAL) The computer system of claim 30 wherein said processor constructs a matrix which associates each word in the documents with a vector and associates each document with a vector.

32. (ORIGINAL) The computer system of claim 31 wherein said processor further decomposes said matrix to define the words and documents as vectors in a continuous vector space.

33. (ORIGINAL) The computer system of claim 31, wherein said processor clusters the files by identifying documents whose vectors are within a threshold distance of one another.

34. (ORIGINAL) The computer system of claim 33, wherein said processor clusters said files in accordance with multiple threshold values to thereby establish plural levels of clusters.

35. (CURRENTLY AMENDED) The computer system of claim 31, wherein said processor automatically labels the clusters based on the resulting clusters.

36. (ORIGINAL) The computer system of claim 35 wherein said processor labels the clusters by selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

37. (NEW) The method according to claim 1, wherein clustering includes organizing the clusters into a hierarchical directory structure.

38. (NEW) A method of organizing a plurality of documents, comprising:

mapping all words of the plurality of documents and the plurality of documents in a semantic vector space;

clustering the plurality of documents to a plurality of clusters based on semantic similarities of the plurality of documents; and

outputting the plurality of documents in a hierarchical format based on a result of clustering the plurality of documents.

39. (NEW) The method according to claim 38, wherein the step of clustering the plurality of documents comprises:

generating the plurality of clusters based on the semantic similarities of the plurality of documents;

organizing the plurality of clusters into one or more granularity levels, wherein the plurality of clusters are organized as directories in the hierarchical format; and

determining a membership of each of the plurality of documents to a particular cluster based on a content of each document in relationship to the particular cluster.

40. (NEW) The method according to claim 39, wherein the step of generating the plurality of clusters comprises:

generating a matrix W of dimensions (M x N), wherein M is a number of a list of words and symbols that occur in the plurality of documents, N is a number of the plurality of documents, each entry $w_{i,j}$ of the matrix W is of the form $w_{i,j} = (1-\varepsilon_i)\frac{c_{i,j}}{n_j}$, $c_{i,j}$ is a number of times a word w_i occurs in a document d_j, n_j is a total number of words in the document d_j, and c_i is a normalized entropy of the word w_i in a corpus T of the collection of the N documents;

performing a singular value decomposition on the matrix W such that $W = USV^T$, wherein U is a (M x R) left singular matrix with row vectors $u_i (1 \le i \le M)$ representing word vectors, S is a diagonal matrix of singular values

$s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is a (N x R) right singular matrix with row vectors $v_i (1 \leq i \leq N)$ representing document vectors, T denotes matrix transposition, and R << M, N;

evaluating closeness of the plurality of documents based on the document vectors; and

merging semantic information of the plurality of documents based on the closeness of the documents.

41. (NEW) The method according to claim 40, wherein the step of evaluating the closeness of the plurality of documents comprises:

measuring closeness of two documents v_j, v_k according to an equation $K(\overline{v_j}, \overline{v_k}) = \cos(v_j S, v_k S) = \frac{v_j S^2 v_k T}{\|v_j S\| \|v_k S\|}$.

42. (NEW) The method according to claim 40, wherein the step of generating the plurality of clusters further comprises:

controlling a number of clusters based on a predetermined threshold of a variance of a cluster resulting from performing the step of merging the semantic information,

wherein the variance σ^2 of the cluster is measured according to an equation $\sigma^2 = \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2}{n_1 + n_2} + \frac{n_1 n_2}{n_1 + n_2} (u_1 + u_2)^2$, u_1, σ_1^2 and u_2, σ_2^2 are means and variances of candidates for merging and n_1 and n_2 are sizes of the clusters.

43. (NEW) The method according to claim 42, wherein the step of generating the plurality of clusters further comprises:

clustering the plurality of documents into the plurality of clusters based on the predetermined threshold variance; and

grouping the plurality of clusters to a one or more super clusters based on one or more additional predetermined threshold variances, wherein super cluster includes one or more lower level clusters and/or super clusters,

wherein each cluster and super cluster is represented as a directory and each super cluster directory is hierarchically above each lower level cluster directory and/or super cluster directory.

44. (NEW) The method according to claim 43, further comprising:

determining whether one or more of the plurality of documents fall outside of the threshold variances of the lowest level of clusters;

determining whether the one or more of the plurality of documents that fall outside of the thresholds of the lowest level of clusters fall within the threshold variance of a particular super cluster; and

grouping the one or more of the plurality of documents to be included in the particular super cluster when it is determined that the one or more of the plurality of documents fall within the threshold variance of the particular super cluster.

45. (NEW) The method according to claim 40, further comprising:

for each cluster, determining a closesness of the words in the documents of the cluster to the cluster; and

choosing one or more words based on the closeness of the of the words to the cluster and assigning the one or more words as a label to the cluster.

46. (NEW) The method according to claim 45, wherein the step of determining the closensess of words and documents in the cluster comprises:

measuring the closeness of a word vector u_i and a document vector v_k according to an equation $\overline{K}(\overline{u_i}, \overline{v_k}) = \cos(u_i S^{1/2}, v_k S^{1/2}) = \frac{u_i S v_k T}{\left\| u_i S^{1/2} \right\| \left\| v_k S^{1/2} \right\|}$.

47. (NEW) The method according to claim 46, wherein the step of choosing the one or more words comprises choosing all words that are within a predetermined closeness threshold *K*.